(19)日本国特許庁（ＪＰ）　(12) **公開特許公報(A)**　(11)特許出願公開番号

特開平7－191599

(43)公開日　平成７年(1995)７月28日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０９Ｂ　21/00	Ｆ			
Ｇ１０Ｌ　3/00	５５１　Ｇ			
	５６１　Ｃ			

審査請求　未請求　請求項の数４　ＯＬ　（全７頁）

(21)出願番号　特願平5－332899

(22)出願日　平成５年(1993)12月27日

(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地

(72)発明者　鳴田　拓生
大阪府門真市大字門真1006番地　松下電器産業株式会社内

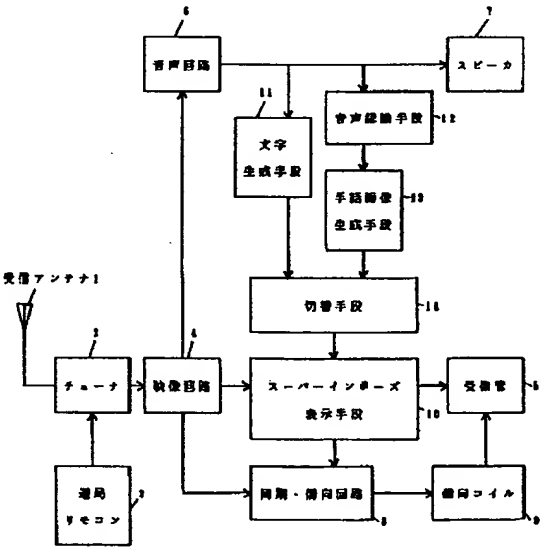
(74)代理人　弁理士　小鍜治　明　（外２名）

(54)【発明の名称】　映像機器

(57)【要約】

【目的】　聾唖者に適したテレビ、ビデオなどの映像機器に関するもので、画面を見るだけで発話内容が理解できることを目的とする。

【構成】　音声信号を文字情報に変換する文字生成手段で生成された文字を画面上に表示させる表示手段を備えた。あるいは音声認識手段で認識された意味内容を手話のアニメーション画像に変換する手話画像生成手段を備えた。これによりテレビ放送をはじめとする映像メディアにおける会話やアナウンスなどの発話内容が文字あるいは手話のアニメーション画像として画面上に即座に表示できる。



1

【特許請求の範囲】

【請求項1】音声信号及び映像信号を入力するメディア入力部と、前記メディア入力部で受けた音声信号を文字情報に変換する文字生成手段と、前記文字生成手段で生成された前記文字情報を画面上に表示させる表示手段とを備えた映像機器。

【請求項2】音声信号及び映像信号を入力するメディア入力部と、前記メディア入力部で受けた音声信号から意味内容を認識する音声認識手段と、前記音声認識手段で認識された意味内容を手話のアニメーション画像に変換する手話画像生成手段と、前記手話画像生成手段で生成された画像情報を画面上に表示させる表示手段とを備えた映像機器。

【請求項3】音声信号及び映像信号を入力するメディア入力部と、前記メディア入力部で受けた音声信号から意味内容を認識する音声認識手段と、前記音声信号ないし映像信号から話者の位置を特定する位置特定手段と、前記音声認識手段で認識された意味内容を手話のアニメーション画像に変換する手話画像生成手段と、前記手話画像生成手段で生成された画像情報を前記位置特定手段で特定された話者の位置に対応して画面上に表示させる表示手段とを備えた映像機器。

【請求項4】音声信号及び映像信号を入力するメディア入力部と、前記メディア入力部で受けた音声信号から意味内容を認識する音声認識手段と、前記音声信号ないし映像信号から話者を識別する話者識別手段と、前記音声認識手段で認識された意味内容を手話のアニメーション画像に変換する手話画像生成手段と、前記手話画像生成手段で生成された画像情報を前記話者識別手段で識別された個々の話者に対応して前記手話のアニメーションの種別を変更し画面上に表示させる表示手段とを備えた映像機器。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は特に聾唖者に適したテレビ、ビデオなどの映像機器に関するものである。

【0002】

【従来の技術】従来のこの種の映像機器の代表例であるテレビは、図8に示すように受信アンテナ1で得た電波をチューナ2に供給し、ここで増幅後周波数変換し選局リモコン3に応じた選局を行う。映像回路4は中間周波を増幅・検波し、音声中間周波と色副搬送波及び同期信号を分配した後、輝度信号の増幅を行い、色差信号を加えて受像管5の電極に所定の映像信号を供給する。音声回路6は音声中間周波を増幅後FM検波・復調し、音声信号に変換してスピーカ7に供給する。同期・偏向回路8は同期信号を分離し、垂直・水平発信器の同期をとり、のこぎり波電流を作成して偏向コイル9に供給する構成となっている。ここで音声信号及び映像信号を入力するメディア入力部は、受信アンテナ1、チューナ2、

2

選曲リモコン3、映像回路4及び音声回路6で構成されている。

【0003】ところで聾唖者のためのテレビ放送としては、番組中の会話やアナウンスなどをテレビ画面上に字幕として表示するクローズドキャプション放送が米国で行われている。日本ではごく限られた番組のみで、音声信号に対応した字幕や手話を付けた映像が放送されている現状である。

【0004】また最近の研究において、不特定話者に対する音声認識や意味情報に対応する手話のアニメーション画像合成が考えられるようになってきた（例えば徐軍、棚橋真、坂本雄児、青木由直："手話画像知的通信のための手振り記述と単語辞書の構成法"信学論A、J76－A、9、pp. 1332－1341（1993－9））。

【0005】さらに有線通信システムの一部には、聾唖者用に音声サービスと同等の内容を文字情報にして画面上にスーパーインポーズ表示させるものがある（例えば特開昭62－48890号公報）。

【0006】

【発明が解決しようとする課題】しかしながら上記従来の構成では、クローズドキャプション放送や字幕、手話付きの放送などのように映像メディアの提供側があらかじめ音声に相当する情報を別途付加しておかない限り、会話やアナウンスなどの発話内容は聾唖者に全く理解できないという課題があった。

【0007】本発明は上記課題を解決するもので、聾唖者などにとって画面を見るだけで発話内容が理解できる映像機器を提供することを目的とする。

【0008】

【課題を解決するための手段】上記課題を解決するために本発明の映像機器は、音声信号及び映像信号を入力するメディア入力部と、このメディア入力部で受けた音声信号を文字情報に変換する文字生成手段と、この文字生成手段で生成された文字情報を画面上に表示させる表示手段とを備えたものである。

【0009】また音声信号及び映像信号を入力するメディア入力部と、このメディア入力部で受けた音声信号から意味内容を認識する音声認識手段と、この音声認識手段で認識された意味内容を手話のアニメーション画像に変換する手話画像生成手段と、この手話画像生成手段で生成された画像情報を画面上に表示させる表示手段とを備えたものである。

【0010】あるいは音声信号及び映像信号を入力するメディア入力部と、このメディア入力部で受けた音声信号から意味内容を認識する音声認識手段と、この音声信号ないし映像信号から話者の位置を特定する位置特定手段と、この音声認識手段で認識された意味内容を手話のアニメーション画像に変換する手話画像生成手段と、この手話画像生成手段で生成された画像情報を位置特定手

段で特定された話者の位置に対応して画面上に表示させる表示手段とを備えたものである。

【0011】あるいは音声信号及び映像信号を入力するメディア入力部と、このメディア入力部で受けた音声信号から意味内容を認識する音声認識手段と、この音声信号ないし映像信号から話者を識別する話者識別手段と、音声認識手段で認識された意味内容を手話のアニメーション画像に変換する手話画像生成手段と、この手話画像生成手段で生成された画像情報を話者識別手段で識別された個々の話者に対応して手話のアニメーションの種別を変更し画面上に表示させる表示手段とを備えたものである。

【0012】

【作用】本発明は上記構成によって、テレビ放送をはじめとするあらゆる映像メディアにおける会話やアナウンスなどの発話内容が文字あるいは手話のアニメーション画像として画面上に即座に表示される。

【0013】特に音声信号ないし映像信号から話者の位置を特定する位置特定手段を備え、この話者の位置に対応して手話のアニメーション画像の表示位置を変えることで、もとの映像信号に合致した臨場感あるわかりやすい画面が構築される。

【0014】あるいは音声信号ないし映像信号から話者を識別する話者識別手段を備え、識別された個々の話者に対応して手話のアニメーションの種別を変更することで、さらに臨場感あるわかりやすい画面が構築される。

【0015】

【実施例】以下本発明の第1の実施例を図1から図3を参照して説明する。図1において従来例で示したものと同一機能を持ち、同一番号を付与し、一部説明を省略する。従来例同様映像機器の代表例であるテレビは、受信アンテナ1で得た電波がまずチューナ2に供給され、ここで増幅後周波数変換し選局リモコン3に応じた選局を行う。映像回路4は中間周波を増幅・検波し、音声中間周波と色副搬送波及び同期信号を分配した後、輝度信号の増幅を行い、色差信号を加えて基本画面に相当する映像信号をスーパーインポーズ表示手段10に供給する。音声回路6は音声中間周波を増幅後FM検波・復調し、音声信号に変換してスピーカ7、文字生成手段11及び音声認識手段12を介し手話画像生成手段13に供給する。ここで音声信号及び映像信号を入力するメディア入力部は、受信アンテナ1、チューナ2、選曲リモコン3、映像回路4及び音声回路6で構成されている。

【0016】文字生成手段11は音声信号を文字情報に変換し、また音声認識手段12は音声信号から意味内容を認識後、手話画像生成手段13で認識された意味内容を手話のアニメーション画像に変換する。切替手段14は文字生成手段11で生成された文字化画像と手話画像生成手段13で生成された手話のアニメーション画像のいずれか一方を選択したり、あるいは両方共を選択したりしなかったりできる切替手段である。切替手段14によって選択された画像はスーパーインポーズ表示手段10で基本画面に相当する映像信号の一部にスーパーインポーズするよう合成された後、受像管5の電極に所定の映像信号を供給する。

【0017】文字生成手段11の構成を図2を用いて説明する。音響分析部11aでは、音声回路6からの音声入力を単位時間ごとのフレーム周期で音響分析し、音声パワーと自己相関関数を求める。この分析データにもとづいて音素認識部11bでは、音声入力を母音と子音に分離し、音素規則部11cにあらかじめ記憶されている音素標準パターンとマッチングをとることによって一音ごとの音素列を抽出する。単語認識部11dでは、単語辞書11eにあらかじめ記憶されている単語知識を用いて単語系列を予測し、これと音素列とのマッチングをとることによって単語を抽出する。構文解析部11fではさらに文法・構文規則部11gにあらかじめ記憶されている言語的知識を用いて文章としての整合性をチェックし修正する。ここで音素規則部11c、単語辞書11e及び文法・構文規則部11gは互いに連携し知識データベースを形成している。漢字かな混じり文字変換部11hでは構文解析部11fで抽出された文章を漢字かな混じりの日本語文字列に変換し、字幕画像生成部11iでこれを画像化して出力する。

【0018】一方音声認識手段12、手話画像生成手段13の構成を図3を用いて説明する。文字生成手段11同様、音響分析部12aでは、音声回路6からの音声入力を単位時間ごとのフレーム周期で音響分析し、音声パワーと自己相関関数を求める。この分析データにもとづいて音素認識部12bでは、音声入力を母音と子音に分離し、音素規則部12cにあらかじめ記憶されている音素標準パターンとマッチングをとることによって一音ごとの音素列を抽出する。単語認識部12dでは、単語辞書12eにあらかじめ記憶されている単語知識を用いて単語系列を予測し、これと音素列とのマッチングをとることによって単語を抽出する。構文解析部12fではさらに文法・構文規則部12gにあらかじめ記憶されている言語的知識を用いて文章としての整合性をチェックし修正する。ここで音素規則部12c、単語辞書12e及び文法・構文規則部12gは互いに連携し知識データベースを形成している。意味推論12hでは構文解析部12fで抽出された文章に対応した意味内容を言語レベルで認識し、基本単位ごとに区切られた意味のある文章情報として手話画像生成手段13に伝送する。

【0019】手話画像生成手段13では、この文章情報に対応して手話記述部13aが手話単語辞書13bにあらかじめ記憶されている手話単語から必要な手話動作パターンを引き出し、合成する。手話画像生成部13cでは手話記述部13aで合成された手話動作をアニメーション画像化して出力する。

【００２０】上記構成において文字生成手段１２が音声信号を文字情報に変換し、さらにスーパーインポーズ表示手段１０がテレビ画面上にこの文字情報を即時に表示するので、スピーカ７からの音声出力がなくてもテレビ画面だけで内容を理解できるという効果がある。あるいは音声認識手段１２、手話画像生成手段１３が音声信号を手話のアニメーション画像に変換し、さらにスーパーインポーズ表示手段１０がテレビ画面上にこの画像情報を即時に表示するので、文字表示だけでは読み取りが困難になるような早口言葉に対しても必要な情報を視覚に訴えてすばやく簡潔に伝えることができる。特に受像管５の画面サイズが小さい場合にも疲れにくいという効果がある。

【００２１】次に本発明の第２の実施例を図４～図５を参照して説明する。図４において本発明の第１の実施例と異なるのは文字生成手段１１や切替手段１４がなく、新たに映像信号及び音声信号から話者の位置を特定する位置特定手段１５とこの話者の位置に対応して画面上に表示させるスーパーインポーズ表示手段１６を備えたことにある。ここで音声回路６からは音声ステレオ信号が左右チャンネル独立に出力されているものとする。

【００２２】他の構成は第１の実施例と同様なので説明を省略し、位置特定手段１５の構成のみ図５を用いて説明する。発話信号抽出部１５ａは、音声回路６からの音声入力から発話信号のみ抽出するフィルタであり、発話信号抽出部１５ａを通過した音声信号について音量検出部１５ｂで音量を検出し、また方位検出部１５ｃで左右の音量バランスとその変化によって音場における話者の方向を算出する。一方、人体位置検出部１５ｄでは、映像回路４からの映像入力を画像処理することで２次元画面から人体位置を検出し、さらに口唇動作検出部１５ｅによって検出された人体中の口唇動作を検出し話者の画面上の位置を算出する。位置判定部１５ｆでは音量検出部１５ｂ、方位検出部１５ｃ及び口唇動作検出部１５ｅからの入力に基づいて仮想の３次元空間における話者の位置を推定し、この位置情報をスーパーインポーズ表示手段１６に伝える。

【００２３】スーパーインポーズ表示手段１６は、位置特定手段１５で特定された話者の位置情報に従ってスーパーインポーズ表示させる手話のアニメーション画像の表示位置を画面上で変更する。スーパーインポーズする２次元表示領域の中で、遠近感はアニメーション画像の大きさによって表現する。

【００２４】上記構成において位置特定手段１５で特定された話者の位置に対応して手話のアニメーション画像の表示位置が即時に変わる。例えば音声信号から判断して右手にいる話者が話をはじめると手話のアニメーションは画面右の方に現れ、話の内容に即した手話動作をはじめる。話者が話しながら左に移動すれば、手話のアニメーションも画面上で左に移動していく。上下方向ない

し奥行き方向に関しても同様である。つまり音声内容を手話に置換しただけでは欠如してしまう話者の位置情報を臨場感を出して画面に盛り込むことができる。

【００２５】次に本発明の第３の実施例を図６～図７を参照して説明する。図６において本発明の第２の実施例と異なるのは、話者の位置を特定する位置特定手段１５がなく、新たに映像信号から話者を識別する話者識別手段１７とこの話者に対応して前記手話のアニメーションの種別を変更し画面上に表示させる表示手段１８を備えたことにある。話者識別手段１７は、受信回路４から入力した映像信号から全ての人を抽出し、これらの人から話者を識別することでアニメーションの種別を変更させる構成である。

【００２６】話者識別手段１７の構成を図７を用いて説明する。発話信号抽出部１７ａは、音声回路６からの音声入力から発話信号のみ抽出するフィルタであり、発話信号抽出部１７ａを通過した音声信号について音量検出部１７ｂ、音程検出部１７ｃ、音色検出部１７ｄ及び速度検出部１７ｅではそれぞれ発話信号の音量、音程、音色及び話者の発語速度を検出する。一方、人体形状検出部１７ｆでは、映像回路４からの映像入力を画像処理することで２次元画面から人体形状を検出し、さらに口唇動作検出部１７ｇ及び特徴抽出部１７ｈによって画像から得られる話者の特徴量を算出する。話者判定部１７ｉではこれら音量検出部１７ｂ、音程検出部１７ｃ、音色検出部１７ｄ、速度検出部１７ｅ、人体形状検出部１７ｆ及び口唇動作検出部１７ｇからの出力に基づき話者の特徴量を例えばあらかじめ用意した２０パターンのいずれかに分類し、この分類された情報をスーパーインポーズ表示手段１８に伝える。分類の方法は主成分分析などの多変量解析手法や学習ベクトル量子化などのニューラルネットワーク手法を用い、多次元空間中に個々人の特徴ベクトルを描くことで実現する。スーパーインポーズ表示手段１８は、話者識別手段１７で識別された個々の話者に対応してスーパーインポーズ表示させる手話のアニメーションの種別を即時に切り替える構成である。

【００２７】上記構成において話者識別手段１７で識別された個々の話者に対応して手話のアニメーションの種別が即時に変わる。つまり音声内容を手話に置換しただけでは欠如してしまう個々の話者の特徴を臨場感を出して画面に盛り込むことができる。

【００２８】なお第１から第３の実施例ではテレビを例にとって説明したが、本発明は音声信号と映像信号を出力するあらゆる映像機器に適応可能である。音声信号や映像信号の種類にも依存しない。画面への表示方式としては、スーパーインポーズ表示手段１０、１６、１８を用いるとしたが、子画面に分割して独立に表示させたり、新たに生成した文字情報や手話のアニメーション画像を別の画面に映し出す構成にしても構わない。画面もＣＲＴなどの受像管５を用いなくてよい。また表示画面

7

を有線ないし無線の通信回線で結び、他の構成要素から離れた場所に設置してもよい。

【0029】さらに文字生成手段11が対象とする言語は日本語に限るものではない。位置特定手段15あるいは話者識別手段17への入力信号は音声信号と映像信号両方を組み合わせなくてもよい。話者識別手段17によって識別される話者の種類は例えば性別など2種類だけでも構わない。逆に話者の特徴量にしたがってアニメーションの大きさ、形状、表情、動作、種別などを無段階に調節しても構わない。声の特徴量を色彩や明度などの色情報に変換し、手話のアニメーション表現に変化を持たせてもよい。

【0030】

【発明の効果】以上のように本発明の映像機器によれば、次の効果が得られる。

【0031】（1）聾唖者や音声なしで映像内容を知りたい人にとって、画面を見るだけで発話内容が理解できる。またこれまでに蓄積された膨大な映像メディアや緊急性の高い報道・中継番組にもそのまま利用できる。

【0032】（2）発話内容を手話のアニメーション画像で表示する場合、文字表示だけでは読み取りが困難になるような早口言葉に対しても必要な情報を視覚に訴えてすばやく簡潔に伝えることができる。手話は、一度習得した人にとっては理解しやすい意志伝達方法であり、表示画面サイズが小さい場合にも疲れにくい効果がある。

【0033】（3）さらに手話のアニメーション画像を特定された話者の位置に対応して画面上に同時に表示させることによって臨場感あるわかりやすい画面が構築できる。

8

【0034】（4）あるいは手話のアニメーション画像の種別を個々の話者に対応して変化させることによってより臨場感あるわかりやすい画面が構築できる。

【図面の簡単な説明】

【図1】本発明の第1の実施例における映像機器の構成図

【図2】同実施例における文字生成手段の構成図

【図3】同実施例における音声認識手段及び手話画像生成手段の構成図

【図4】本発明の第2の実施例における映像機器の構成図

【図5】同実施例における位置特定手段の構成図

【図6】本発明の第3の実施例における映像機器の構成図

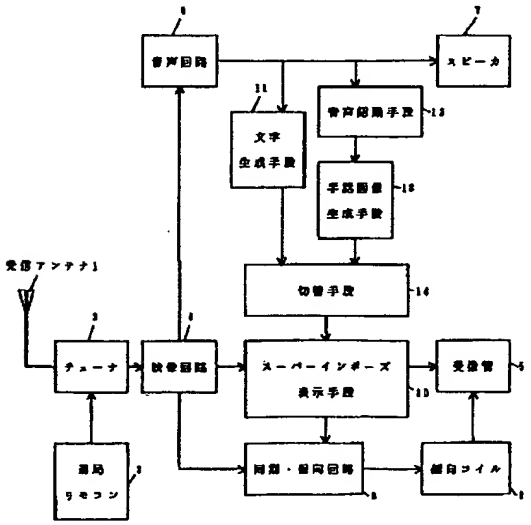
【図7】同実施例における話者識別手段の構成図

【図8】従来の映像機器の構成図

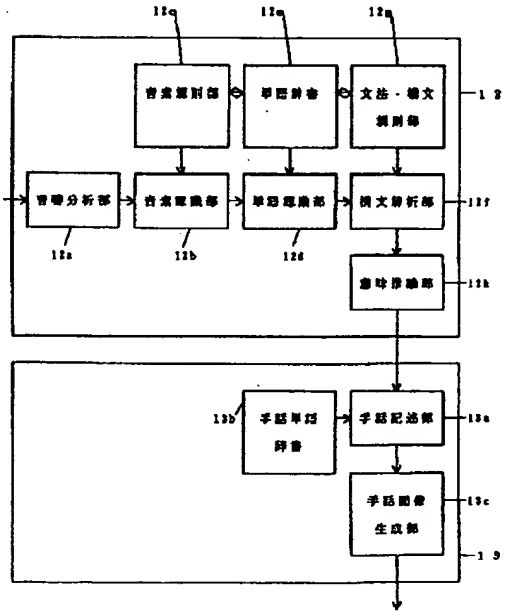
【符号の説明】

2　チューナ
4　映像回路
6　音声回路
10　スーパーインポーズ表示手段
11　文字生成手段
12　音声認識手段
13　手話画像生成手段
15　位置特定手段
17　話者識別手段

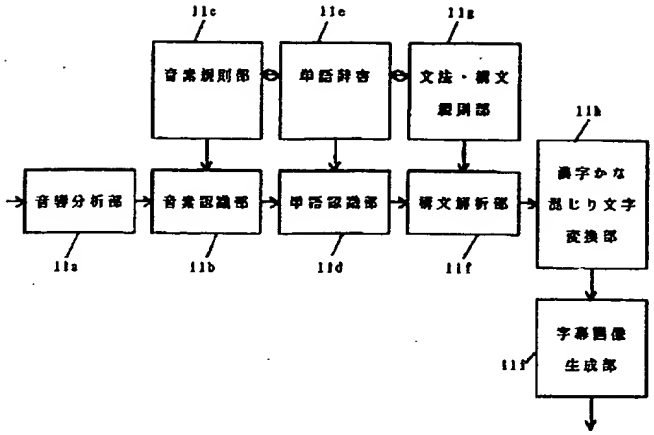
【図1】



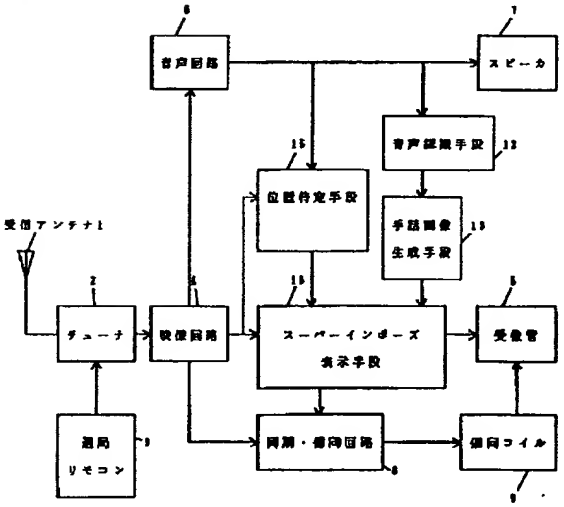
【図3】



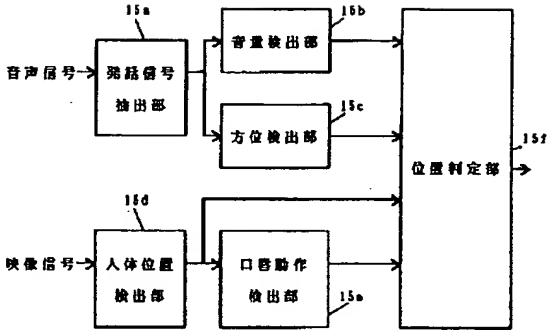
【図2】


音響分析部
音素認識部
単語認識部
構文解析部
漢字かな混じり文字変換部
字幕画像生成部
音素規則部
単語辞書
文法・構文規則部

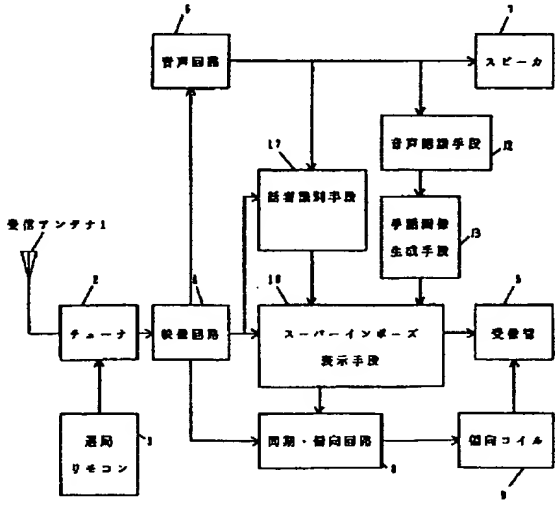
【図4】


音声回路
スピーカ
音声認識手段
位置特定手段
手話画像生成手段
受信アンテナ1
チューナ
映像回路
スーパーインポーズ表示手段
受像管
選局リモコン
同期・偏向回路
偏向コイル

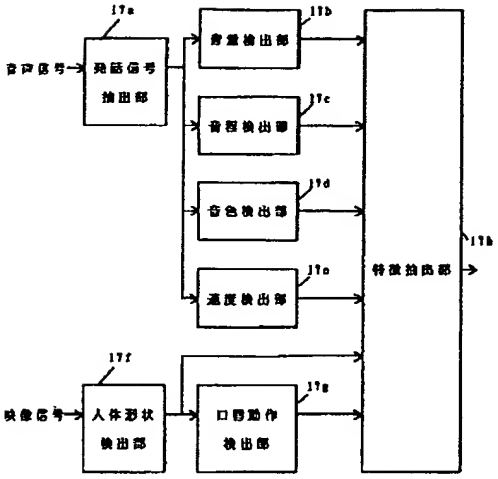
【図5】


音声信号
発話信号抽出部
音量検出部
方位検出部
位置判定部
映像信号
人体位置検出部
口唇動作検出部

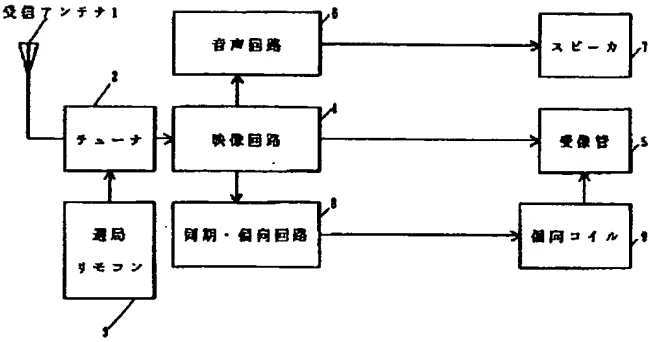
【図6】


音声回路
スピーカ
音声認識手段
話者識別手段
手話画像生成手段
受信アンテナ1
チューナ
映像回路
スーパーインポーズ表示手段
受像管
選局リモコン
同期・偏向回路
偏向コイル

【図7】


音声信号
発話信号抽出部
音量検出部
音程検出部
音色検出部
速度検出部
映像信号
人体形状検出部
口唇動作検出部
特徴抽出部

【図8】


受信アンテナ1
チューナ
選局リモコン
音声回路
映像回路
同期・偏向回路
スピーカ
受像管
偏向コイル

B1

CLAIMS

[Claim(s)]
[Claim 1] The visual equipment equipped with the media input section which inputs a sound signal and a video signal, a character generation means to change into alphabetic information the sound signal received in the aforementioned media input section, and a display means to display on a screen the aforementioned alphabetic information generated with the aforementioned character generation means.
[Claim 2] The visual equipment equipped with a speech recognition means to recognize a semantic content from the sound signal received in the media input section which inputs a sound signal and a video signal, and the aforementioned media input section, a sign language picture generation means to change into the animation picture of sign language the semantic content recognized with the aforementioned speech recognition means, and a display means to display on a screen the image information generated with the aforementioned sign language picture generation means.
[Claim 3] The visual equipment characterized by providing the following. The media input section which inputs a sound signal and a video signal. A speech recognition means to recognize a semantic content from the sound signal received in the aforementioned media input section. A position specification means to pinpoint a speaker's position from the aforementioned sound signal or a video signal. A sign language picture generation means to change into the animation picture of sign language the semantic content recognized with the aforementioned speech recognition means, and a display means to display on a screen the image information generated with the aforementioned sign language picture generation means corresponding to a speaker's position pinpointed with the aforementioned position specification means.
[Claim 4] The visual equipment characterized by providing the following. The media input section which inputs a sound signal and a video signal. A speech recognition means to recognize a semantic content from the sound signal received in the aforementioned media input section. A speaker identification means to discriminate a speaker from the aforementioned sound signal or a video signal. The display means which changes the classification of the animation of the aforementioned sign language corresponding to each speaker discriminated with the aforementioned speaker identification means in the image information generated with a sign language picture generation means to change into the animation picture of sign language the semantic content recognized with the aforementioned speech recognition means, and the aforementioned sign language picture generation means, and is displayed on a screen.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]

[Industrial Application] Especially this invention relates to visual equipments suitable for the deaf-mute, such as television and video.
[0002]
[Description of the Prior Art] Television which is the example of representation of this conventional kind of visual equipment supplies the electric wave acquired by the receiving antenna 1 to a tuner 2, as shown in drawing 8 , and after [amplification] frequency conversion of it is carried out here, and it performs the channel selection according to the channel selection remote control 3. After the image circuit 4 amplifies and detects an intermediate frequency and distributes a voice intermediate frequency, a chrominance subcarrier, and a synchronizing signal, it performs amplification of a luminance signal, adds a color-difference signal, and supplies a predetermined video signal to the electrode of the picture tube 5. It after [amplification]-FM-detects, and the voice circuit 6 restores to a voice intermediate frequency, changes it into a sound signal, and is supplied to a loudspeaker 7. A synchronization and a deflection circuit 8 separate a synchronizing signal, takes the synchronization of a perpendicular and a level transmitter, and has composition which creates sawtooth wave current and is supplied to a deflecting coil 9. The media input section which inputs a sound signal and a video signal here consists of a receiving antenna 1, a tuner 2, song selection remote control 3, an image circuit 4, and a voice circuit 6.
[0003] By the way, as television broadcasting for a deaf-mute, closed caption broadcast which displays conversation, an announcement, etc. in a program as a title on a television screen is performed in the U.S. In Japan, it is only the program restricted very much and is the present condition that the image which attached the title and sign language corresponding to the sound signal is broadcast.
[0004] Moreover, in the latest research, animation picture composition of the sign language corresponding to the speech recognition and the semantic information over a speaker independence can be considered now (for example, construction "******A of the gesture description for ****, Tanahashi truth, Takeji Sakamoto, and Yoshinao Aoki:" sign language picture intellectual communication, and a word dictionary, J76-A, 9, pp.1332-1341 (1993-9)).
[0005] Furthermore, some wire communication systems have some which a content equivalent [for deaf-mutes] to voice service is made [some] into alphabetic information, and indicate by superimposition on a screen (for example, JP,62-48890,A).
[0006]
[Problem(s) to be Solved by the Invention] However, with the above-mentioned conventional composition, unless the offer side of image media added separately the information which is equivalent to voice beforehand like closed caption broadcast or broadcast with a title and sign language, the contents of utterance, such as conversation and an announcement, had the technical problem that he could not understand at all to the deaf-mute.
[0007] this invention solves the above-mentioned technical problem, and it aims at offering the visual equipment which can understand the content of utterance only by seeing a screen for a deaf-mute etc.
[0008]
[Means for Solving the Problem] In order to solve the above-mentioned technical problem,

the visual equipment of this invention is equipped with the media input section which inputs a sound signal and a video signal, a character generation means to change into alphabetic information the sound signal received in this media input section, and a display means to display on a screen the alphabetic information generated with this character generation means.

[0009] Moreover, it has a speech recognition means to recognize a semantic content from the sound signal received in the media input section which inputs a sound signal and a video signal, and this media input section, a sign language picture generation means to change into the animation picture of sign language the semantic content recognized with this speech recognition means, and a display means to display on a screen the image information generated with this sign language picture generation means.

[0010] Or the media input section which inputs a sound signal and a video signal and a speech recognition means to recognize a semantic content from the sound signal received in this media input section, A position specification means to pinpoint a speaker's position from this sound signal or a video signal, It has a sign language picture generation means to change into the animation picture of sign language the semantic content recognized with this speech recognition means, and a display means to display on a screen the image information generated with this sign language picture generation means corresponding to a speaker's position pinpointed with the position specification means.

[0011] Or the media input section which inputs a sound signal and a video signal and a speech recognition means to recognize a semantic content from the sound signal received in this media input section, A speaker identification means to discriminate a speaker from this sound signal or a video signal, and a sign language picture generation means to change into the animation picture of sign language the semantic content recognized with the speech recognition means, It has the display means which changes the classification of the animation of sign language corresponding to each speaker discriminated with the speaker identification means in the image information generated with this sign language picture generation means, and is displayed on a screen.

[0012]

[Function] As for this invention, the contents of utterance, such as conversation, an announcement, etc. in all image media, including television broadcasting are displayed immediately on a screen by the above-mentioned composition as an animation picture of a character or sign language.

[0013] It has a position specification means to pinpoint a speaker's position from a sound signal or a video signal especially, and the intelligible screen corresponding to the video signal of a basis which exists a feeling of presence is built by changing the display position of the animation picture of sign language corresponding to this speaker's position.

[0014] Or the intelligible screen which exists a feeling of presence further is built by having a speaker identification means to discriminate a speaker from a sound signal or a video signal, and changing the classification of the animation of sign language corresponding to each discriminated speaker.

[0015]

[Example] The 1st example of this invention is explained with reference to drawing 3 from drawing 1 below. It has the same function as what was shown in the conventional example

in drawing 1 , the same number is given, and explanation is omitted in part. The electric wave acquired by the receiving antenna 1 is first supplied to a tuner 2, after [amplification] frequency conversion of the television which is the example of representation of a visual equipment as well as the conventional example is carried out here, and it performs the channel selection according to the channel selection remote control 3. After the image circuit 4 amplifies and detects an intermediate frequency and distributes a voice intermediate frequency, a chrominance subcarrier, and a synchronizing signal, it performs amplification of a luminance signal and supplies the video signal which adds a color-difference signal and is equivalent to a basic screen to the superimposition display means 10. It after [amplification]-FM-detects, and the voice circuit 6 restores to a voice intermediate frequency, changes it into a sound signal, and is supplied to the sign language picture generation means 13 through a loudspeaker 7, the character generation means 11, and the speech recognition means 12. The media input section which inputs a sound signal and a video signal here consists of a receiving antenna 1, a tuner 2, song selection remote control 3, an image circuit 4, and a voice circuit 6.

[0016] The character generation means 11 changes a sound signal into alphabetic information, and the speech recognition means 12 changes into the animation picture of sign language the semantic content recognized with the sign language picture generation means 13 after recognizing a semantic content from the sound signal. The change means 14 is a change means which cannot choose the character-ized picture generated with the character generation means 11, or the animation picture of the sign language generated with the sign language picture generation means 13, or cannot choose both. The picture chosen by the change means 14 supplies a predetermined video signal to the electrode of the picture tube 5, after being compounded so that it may superimpose to a part of video signal which is equivalent to a basic screen with the superimposition display means 10.

[0017] The composition of the character generation means 11 is explained using drawing 2 . In acoustic-analysis section 11a, acoustic analysis of the voice input from the voice circuit 6 is carried out by the frame period for every unit time, and voice power and an autocorrelation function are searched for. Based on these analytical data, by phoneme recognition section 11b, voice input is divided into a vowel and a consonant and the phoneme train for every sound is extracted by taking the phoneme standard pattern and matching which are beforehand memorized by phoneme rule section 11c. In 11d of word recognition sections, a word sequence is predicted using the word knowledge beforehand memorized by word dictionary 11e, and a word is extracted by taking matching with this and a phoneme train. In 11f of syntax analyzers, the adjustment as a text is checked and modified to syntax and 11g of grammatical rule sections using the linguistic knowledge memorized beforehand. Phoneme rule section 11c, word dictionary 11e, and 11g of syntax and the grammatical rule sections cooperate mutually, and they form the knowledge database here. In 11h of sentence mixing kanji, kana and characters character transducers, the text extracted by 11f of syntax analyzers is changed into the Japanese character string of kanji kana mixture, and this is imaged and outputted by title picture generation section 11i.

[0018] On the other hand, the composition of the speech recognition means 12 and the sign language picture generation means 13 is explained using drawing 3 . Like the character generation means 11, by acoustic-analysis section 12a, acoustic analysis of the voice input

from the voice circuit 6 is carried out by the frame period for every unit time, and voice power and an autocorrelation function are searched for. Based on these analytical data, by phoneme recognition section 12b, voice input is divided into a vowel and a consonant and the phoneme train for every sound is extracted by taking the phoneme standard pattern and matching which are beforehand memorized by phoneme rule section 12c. In 12d of word recognition sections, a word sequence is predicted using the word knowledge beforehand memorized by word dictionary 12e, and a word is extracted by taking matching with this and a phoneme train. In 12f of syntax analyzers, the adjustment as a text is checked and modified to syntax and 12g of grammatical rule sections using the linguistic knowledge memorized beforehand. Phoneme rule section 12c, word dictionary 12e, and 12g of syntax and the grammatical rule sections cooperate mutually, and they form the knowledge database here. In 12h of semantic reasoning, the semantic content corresponding to the text extracted by 12f of syntax analyzers is recognized on language level, and it transmits to the sign language picture generation means 13 as text information with the meaning divided for every base unit.

[0019] With the sign language picture generation means 13, sign language description section 13a pulls out and compounds a required sign language operation pattern corresponding to this text information from the sign language word beforehand memorized by sign language word dictionary 13b. In sign language picture generation section 13c, animation imaging is carried out and sign language operation compounded by sign language description section 13a is outputted.

[0020] Since the character generation means 12 changes a sound signal into alphabetic information in the above-mentioned composition and the superimposition display means 10 displays this alphabetic information immediately on a television screen further, even if there is no voice output from a loudspeaker 7, it is effective in the ability to understand the content only on a television screen. Or it can complain of required information to a visual sense also to a tongue twister to which reading becomes difficult only by character representation since the speech recognition means 12 and the sign language picture generation means 13 change a sound signal into the animation picture of sign language and the superimposition display means 10 displays this image information immediately on a television screen further, and can tell quickly and briefly. It is effective in being hard to get tired, when especially the screen size of the picture tube 5 is small.

[0021] Next, the 2nd example of this invention is explained with reference to drawing 4 - drawing 5 . There is no differing [of the character generation means 11 or 14 change means] from the 1st example of this invention in drawing 4 , and it is in having had the superimposition display means 16 displayed on a screen corresponding to the position of a position specification means 15 to newly pinpoint a speaker's position from a video signal and a sound signal, and this speaker. From the voice circuit 6, the voice stereo signal shall be outputted to right-and-left channel independence here.

[0022] Since other composition is the same as that of the 1st example, it omits explanation, and it explains only the composition of the position specification means 15 using drawing 5 . Utterance signal extraction section 15a is a filter which extracts only an utterance signal from the voice input from the voice circuit 6, detects volume by volume detecting-element 15b about the sound signal which passed utterance signal extraction section 15a, and

computes the direction of a speaker in sound field by volume balance and its change on either side by direction detecting-element 15c. On the other hand, in 15d of human body position detecting elements, labial operation in the human body which detected the human body position from the two-dimensional screen by carrying out the image processing of the image input from the image circuit 4, and was further detected by labial operation detecting-element 15e is detected, and the position on a speaker's screen is computed. In 15f of position judging sections, based on the input from volume detecting-element 15b, direction detecting-element 15c, and labial operation detecting-element 15e, a speaker's position in the 3-dimensional space of imagination is presumed, and this positional information is told to the superimposition display means 16.

[0023] The superimposition display means 16 changes on a screen the display position of the animation picture of sign language which indicates by superimposition according to a speaker's positional information specified with the position specification means 15. In the two-dimensional viewing area to superimpose, the size of an animation picture expresses depth perception.

[0024] Corresponding to a speaker's position pinpointed with the position specification means 15 in the above-mentioned composition, the display position of the animation picture of sign language changes immediately. For example, if the speaker who is in a right hand judging from a sound signal begins the talk, the animation of sign language will appear in the direction of the screen right, and will begin sign language operation adapted to the content of the talk. If it moves to the left while a speaker talks, the animation of sign language also moves to the left on the screen. It is the same also about the vertical direction or the depth direction. That is, presence can be taken out and a speaker's positional information lacked only by replacing the content of voice by sign language can be incorporated in a screen.

[0025] Next, the 3rd example of this invention is explained with reference to drawing 6 - drawing 7 . It is to differ from the 2nd example of this invention in drawing 6 to have had the display means 18 which there is no position specification means 15 to pinpoint a speaker's position, changes the classification of the animation of the aforementioned sign language corresponding to a speaker identification means 17 to newly discriminate a speaker from a video signal, and this speaker, and is displayed on a screen. The speaker identification means 17 is the composition of making the classification of animation changing by extracting all men from the video signal inputted from the receiving circuit 4, and discriminating a speaker from these men.

[0026] The composition of the speaker identification means 17 is explained using drawing 7 . Utterance signal extraction section 17a is a filter which extracts only an utterance signal from the voice input from the voice circuit 6, and detects the volume of an utterance signal, a pitch, a tone, and a speaker's speech speed about the sound signal which passed utterance signal extraction section 17a, respectively by volume detecting-element 17b, pitch detecting-element 17c, 17d of tone detecting elements, and speed-detector 17e. On the other hand, in 17f of human body configuration detecting elements, a human body configuration is detected from a two-dimensional screen by carrying out the image processing of the image input from the image circuit 4, and a speaker's characteristic quantity further obtained from a picture by 17g of labial operation detecting elements and

17h of feature-extraction sections is computed. It classifies into either of 20 patterns which prepared a speaker's characteristic quantity beforehand based on the output from these volume detecting-element 17b, pitch detecting-element 17c, 17d of tone detecting elements, speed-detector 17e, 17f of human body configuration detecting elements, and 17g of labial operation detecting elements, for example according to speaker judging section 17i, and this classified information is told to the superimposition display means 18. The method of a classification is realized by drawing an individual feature vector all over multi-dimension space using the neural network technique, such as the multivariate-analysis technique, such as principal component analysis, and a study vector quantization. The superimposition display means 18 is composition which changes immediately the classification of the animation of sign language which indicates by superimposition corresponding to each speaker discriminated with the speaker identification means 17.

[0027] Corresponding to each speaker discriminated with the speaker identification means 17 in the above-mentioned composition, the classification of the animation of sign language changes immediately. That is, presence can be taken out and each speaker's feature lacked only by replacing the content of voice by sign language can be incorporated in a screen.

[0028] In addition, although the 1st to 3rd example explained taking the case of television, this invention can be adapted for all the visual equipments that output a sound signal and a video signal. It is not dependent on the kind of a sound signal or video signal. As means of displaying to a screen although the superimposition display meanses 10, 16, and 18 are used, it divides into a child screen, and it may be made to display independently or the alphabetic information and the animation picture of sign language which were newly generated may be made the composition projected on another screen. A screen does not need to use the picture tubes 5, such as CRT,, either. Moreover, the display screen may be connected with the communication line of a cable or radio, and you may install in the place distant from other components.

[0029] Furthermore, the character generation means 11 does not restrict the target language to Japanese. The input signal to the position specification means 15 or the speaker identification means 17 does not need to combine both a sound signal and a video signal. Two kinds of sex etc. does not care about speaker's kind discriminated by the speaker identification means 17. Conversely, according to a speaker's characteristic quantity, you may adjust the size of animation, a configuration, expression, operation, classification, etc. on a stepless story. The characteristic quantity of voice may be changed into sexual desire news, such as color and lightness, and change may be given to animation expression of sign language.

[0030]

[Effect of the Invention] According to the visual equipment of this invention, the following effect is acquired as mentioned above.

[0031] (1) For those who want to know the content of an image without a deaf-mute or voice, he can understand the content of utterance only by seeing a screen. Moreover, it can use also for high report / relay program of huge image media or urgency accumulated until now as it is.

[0032] (2) When displaying the content of utterance by the animation picture of sign language, only by character representation, also to a tongue twister to which reading

becomes difficult, it can complain of required information to a visual sense, and it can be told briefly quickly. For the person who mastered at once, sign language is the volition transfer method which is easy to understand, and is effective in being hard to get tired, when display screen size is small.

[0033] (3) The intelligible screen which exists a feeling of presence can be built by making it display simultaneously on a screen corresponding to the position of the speaker who had the animation picture of sign language specified further.

[0034] (4) or the thing for which the classification of the animation picture of sign language is changed corresponding to each speaker -- more -- a feeling of presence -- a certain intelligible screen can be built

---

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] The block diagram of the visual equipment in the 1st example of this invention

[Drawing 2] The block diagram of the character generation means in this example

[Drawing 3] The block diagram of the speech recognition means in this example, and a sign language picture generation means

[Drawing 4] The block diagram of the visual equipment in the 2nd example of this invention

[Drawing 5] The block diagram of the position specification means in this example

[Drawing 6] The block diagram of the visual equipment in the 3rd example of this invention

[Drawing 7] The block diagram of the speaker identification means in this example

[Drawing 8] The block diagram of the conventional visual equipment

[Description of Notations]

2 Tuner
4 Image Circuit
6 Voice Circuit
10 Superimposition Display Means
11 Character Generation Means
12 Speech Recognition Means
13 Sign Language Picture Generation Means
15 Position Specification Means
17 Speaker Identification Means